(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2018-149629
(P2018-149629A)
(43)公開日　平成30年9月27日(2018.9.27)

(51) Int.Cl.		FI			テーマコード（参考）
B25J 13/00	(2006.01)	Ｂ２５Ｊ	13/00	Ｚ	２Ｃ１５０
B25J 13/08	(2006.01)	Ｂ２５Ｊ	13/08	Ｚ	３Ｃ７０７
G10L 15/10	(2006.01)	Ｇ１０Ｌ	15/10	５００Ｎ	
G10L 25/63	(2013.01)	Ｇ１０Ｌ	25/63		
G10L 25/15	(2013.01)	Ｇ１０Ｌ	25/15		

審査請求 未請求　請求項の数 11　ＯＬ　（全 17 頁）　最終頁に続く

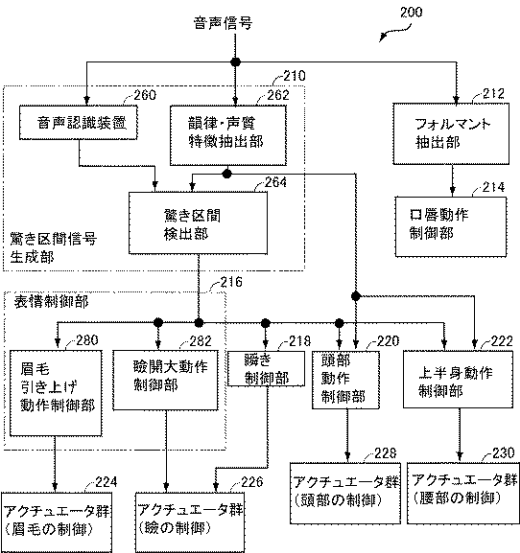
(21)出願番号　特願2017-47509(P2017-47509)
(22)出願日　平成29年3月13日(2017.3.13)

（出願人による申告）平成２６－２８年度　国立研究開発法人科学技術振興機構　戦略的創造研究推進事業　総括実施型研究ＥＲＡＴＯ「石黒共生ヒューマンロボットインタラクションプロジェクト」に係る委託業務、産業技術力強化法第１９条の適用を受ける特許出願

(71)出願人　393031586
株式会社国際電気通信基礎技術研究所
京都府相楽郡精華町光台二丁目２番地２
(74)代理人　100099933
弁理士　清水　敏
(72)発明者　イシイ　カルロス　トシノリ
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内
(72)発明者　港　隆史
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内

Ｆターム（参考）　2C150 AA05 CA01 DA25 DA27 DF03
DF33 ED42 EF29
3C707 AS36 KS11 KS39 MT14 WA03
WL07 WL15 WL16 WM25

(54)【発明の名称】ヒューマノイドロボットの驚き動作生成装置

(57)【要約】
【課題】ヒューマノイドロボットにおいて、驚きに伴う、自然な、人らしい動作を実現する。
【解決手段】驚き動作生成装置２００は、驚き区間信号を受信し、開始時刻に、表情を驚き状態に変化させる動作を開始し、終了時刻の後、第１の戻り時間内に表情を中立状態に戻すようロボットを制御する表情制御部２１６と、開始時刻に頭部を驚き状態に変化させる動作を開始し、終了時刻の後、第２の戻り時間内に、頭部の状態を中立状態に戻すようロボットを制御する頭部動作制御部２２０と、開始時刻に上半身を驚き状態に変化させる動作を開始し、終了時刻の後、第３の戻り時間内に上半身を中立状態に戻すようロボットを制御する上半身動作制御部２２２とを含む。これらのいずれにおいても、戻り時間が、驚き状態への変化に要する時間よりも長くなるように制御される。
【選択図】図１２

【特許請求の範囲】
【請求項１】
所定の時間区間において、驚きに伴う動作を行う様にヒューマノイド型のロボットを制御するための驚き動作生成装置であって、
前記ロボットの顔の表情、頭部の状態、及び上半身の状態は、驚いていない状態である中立状態と、驚いた状態である驚き状態とにそれぞれ制御可能であり、
前記時間区間の少なくとも開始時刻と終了時刻とを特定する驚き区間信号を受信する様に接続され、前記開始時刻に、前記顔の表情を前記驚き状態に変化させる動作を開始し、前記終了時刻の後、第１の戻り時間内に、前記顔の表情を前記中立状態に戻すよう、前記ロボットを制御する表情制御手段と、
前記開始時刻に、前記頭部を前記驚き状態に変化させる動作を開始し、前記終了時刻の後、第２の戻り時間内に、前記頭部の状態を前記中立状態に戻すよう、前記ロボットを制御する頭部制御手段と、
前記開始時刻に、前記上半身を前記驚き状態に変化させる動作を開始し、前記終了時刻の後、第３の戻り時間内に、前記上半身を前記中立状態に戻すよう、前記ロボットを制御する上半身制御手段とを含み、
前記表情制御手段、前記頭部制御手段、及び前記上半身制御手段のいずれにおいても、前記第１、第２又は第３の戻り時間が、前記開始時刻に開始した、前記驚き状態への変化に要する時間よりも長くなる様に制御される、驚き動作生成装置。
【請求項２】
前記驚き状態は、第１のレベルの驚き状態と、前記第１のレベルより高い第２のレベルの驚き状態とを含み、
前記驚き区間信号に加えて、驚きのレベルを示す驚きレベル信号がさらに前記驚き動作生成装置に与えられ、
前記表情制御手段、前記頭部制御手段、及び前記上半身制御手段の任意の組み合わせにおいて、前記驚きレベル信号に応じて驚き状態を前記第１のレベルに応じた第１の驚き状態と、前記第２のレベルに応じた第２の驚き状態とに区別して前記ロボットを制御する、請求項１に記載の驚き動作生成装置。
【請求項３】
前記中立状態から前記第２の驚き状態への前記ロボットの各部の変位量は、前記中立状態から前記第１の驚き状態への前記ロボットの各部の変位量より大きい、請求項２に記載の驚き動作生成装置。
【請求項４】
前記表情制御手段は、前記驚き状態における眉毛の目からの離間量が、前記中立状態における離間量より大きくなる様に眉毛を制御する、請求項１～請求項３のいずれかに記載の驚き動作生成装置。
【請求項５】
前記頭部制御手段は、前記驚き状態における前記頭部が、前記中立状態における位置から前記ロボットにとって後方に移動する様に前記頭部を制御する、請求項１～請求項４のいずれかに記載の驚き動作生成装置。
【請求項６】
前記上半身制御手段は、前記驚き状態における前記上半身が、前記中立状態における位置から前記ロボットにとって後方に反る様に前記上半身を制御する、請求項１～請求項５のいずれかに記載の驚き動作生成装置。
【請求項７】
前記表情制御手段、前記頭部制御手段、及び前記上半身制御手段のいずれも、前記中立状態と前記驚き状態との間の変化が滑らかになる様に前記ロボットを制御する、請求１～請求項６のいずれかに記載の驚き動作生成装置。
【請求項８】
前記顔の表情が、前記驚き状態から前記中立状態に戻ったことに応答して、前記ロボット

の目を１回閉じさせた後、前記中立状態に戻す制御を行うことにより前記ロボットに瞬きを行わせるための瞬き制御手段をさらに含む、請求項１～請求項７のいずれかに記載の驚き動作生成装置。
【請求項９】
前記時間区間において、前記ロボットが発声すべき発話音声信号のフォルマントを抽出するフォルマント抽出手段と、
前記フォルマント抽出手段により抽出されたフォルマントに対応して前記ロボットの口唇の開口量を制御するための口唇動作制御手段をさらに含む、請求項１～請求項８のいずれかに記載の驚き動作生成装置。
【請求項１０】
音声信号を受信し、当該音声信号から、その話者の驚き状態を検出し、前記驚き区間信号を生成するための驚き区間信号生成手段をさらに含む、請求項１～請求項９のいずれかに記載の驚き動作生成装置。
【請求項１１】
前記音声信号からフォルマントを抽出するフォルマント抽出手段と、
前記フォルマント抽出手段により抽出されたフォルマントに対応して前記ロボットの口唇の開口量を制御するための口唇動作制御手段をさらに含む、請求項１０に記載の驚き動作生成装置。
【発明の詳細な説明】
【技術分野】
【０００１】
この発明は、いわゆるヒューマノイドロボットの制御の改善技術に関し、特に、人間がヒューマノイドロボットを相手に対話しているときのヒューマノイドロボットの動作を従来のものより人に近い自然なものにする技術に関する。
【背景技術】
【０００２】
驚き発話は日常会話の場面で頻出し、コミュニケーションを円滑に進める上で感情や態度の表現において重要な役割を果たす。
【０００３】
これまで人型ロボットを通して円滑な遠隔コミュニケーションを実現する研究が行われてきたが、人型のヒューマノイドロボット、特に人の姿を持つアンドロイドの場合は、より人らしい振る舞いが要求されることが明らかになっている。特に、驚きのような感情表現において、声は驚いているのに適切な表情又は動作が伴わない場合は、コミュニケーションの相手に違和感が生じる。したがって、驚き発話に伴う自然な動作を生成することは重要である。
【０００４】
過去には、笑いに伴うアンドロイドの動作について自然な動作を実現することを目指した技術を開示したもの（後掲の非特許文献１）はあるが、対話において生じる驚きに伴う自然な動作をアンドロイドに行わせるような技術は実現できていない。
【先行技術文献】
【非特許文献】
【０００５】
【非特許文献１】Ishi, C., Funayama, T., Minato, T., Ishiguro, H. (2016). “Motion generation in android robots during laughing speech,” Proc. IROS 2016, pp. 3327-3332, Oct., 2016.
【発明の概要】
【発明が解決しようとする課題】
【０００６】
既に述べたとおり、アンドロイドの様に人間によく似たロボットの場合、対話中に人間と異なる動作をすると違和感が生じ、かえって不気味な印象を相手に与えてしまう。その

ような問題を生じさせないためには、対話中のアンドロイドの動作をごく自然なものにする必要がある。特に、驚き動作は、対話を滑らかに進行させる上で非常に重要であるにもかかわらず、そうした問題に着目してアンドロイドの動作を制御する技術はこれまで存在していなかった。

【０００７】

したがって、ヒューマノイドロボットにおいて、驚きに伴う、自然な、人らしい動作を実現するための制御装置を提供することが本発明の１つの目標である。

【０００８】

また、驚きに伴う、自然な、人らしい動作を実現するためにヒューマノイドロボットの各部を協調して動作させる制御装置を提供することも好ましい。

【課題を解決するための手段】

【０００９】

第１の局面に係る驚き動作生成装置は、所定の時間区間において、驚きに伴う動作を行う様にヒューマノイド型のロボットを制御する。ロボットの顔の表情、頭部の状態、及び上半身の状態は、驚いていない状態である中立状態と、驚いた状態である驚き状態とにそれぞれ制御可能である。この驚き動作生成装置は、上記時間区間の少なくとも開始時刻と終了時刻とを特定する驚き区間信号を受信する様に接続され、開始時刻に、顔の表情を驚き状態に変化させる動作を開始し、終了時刻の後、第１の戻り時間内に、顔の表情を中立状態に戻すよう、ロボットを制御する表情制御手段と、開始時刻に、頭部を驚き状態に変化させる動作を開始し、終了時刻の後、第２の戻り時間内に、頭部の状態を中立状態に戻すよう、ロボットを制御する頭部制御手段と、開始時刻に、上半身を驚き状態に変化させる動作を開始し、終了時刻の後、第３の戻り時間内に、上半身を中立状態に戻すよう、ロボットを制御する上半身制御手段とを含む。表情制御手段、頭部制御手段、及び上半身制御手段のいずれにおいても、第１、第２又は第３の戻り時間が、開始時刻に開始した、驚き状態への変化に要する時間よりも長くなる様に制御される。

【００１０】

好ましくは、驚き状態は、第１のレベルの驚き状態と、第１のレベルより高い第２のレベルの驚き状態とを含む。また、驚き区間信号に加えて、驚きのレベルを示す驚きレベル信号がさらに驚き動作生成装置に与えられる。表情制御手段、頭部制御手段、及び上半身制御手段の任意の組み合わせにおいて、驚きレベル信号に応じて驚き状態を第１のレベルに応じた第１の驚き状態と、第２のレベルに応じた第２の驚き状態とに区別してロボットを制御する。

【００１１】

より好ましくは、中立状態から第２の驚き状態へのロボットの各部の変位量は、中立状態から第１の驚き状態へのロボットの各部の変位量より大きい。

【００１２】

さらに好ましくは、表情制御手段は、驚き状態における眉毛の目からの離間量が、中立状態における離間量より大きくなる様に眉毛を制御する。

【００１３】

頭部制御手段は、驚き状態における頭部が、中立状態における位置からロボットにとって後方に移動する様に頭部を制御してもよい。

【００１４】

好ましくは、上半身制御手段は、驚き状態における上半身が、中立状態における位置からロボットにとって後方に反る様に上半身を制御する。

【００１５】

より好ましくは、表情制御手段、頭部制御手段、及び上半身制御手段のいずれも、中立状態と驚き状態との間の変化が滑らかになる様にロボットを制御する。

【００１６】

さらに好ましくは、驚き動作生成装置は、ロボットの顔の表情が、驚き状態から中立状態に戻ったことに応答して、ロボットの目を１回閉じさせた後、中立状態に戻す制御を行

うことによりロボットに瞬きを行わせるための瞬き制御手段をさらに含む。
【００１７】
　驚き動作生成装置は、時間区間において、ロボットが発声すべき発話音声信号のフォルマントを抽出するフォルマント抽出手段と、フォルマント抽出手段により抽出されたフォルマントに対応してロボットの口唇の開口量を制御するための口唇動作制御手段をさらに含んでもよい。
【００１８】
　好ましくは、驚き動作生成装置は、音声信号を受信し、当該音声信号から、その話者の驚き状態を検出し、驚き区間信号を生成するための驚き区間信号生成手段をさらに含む。
【００１９】
　より好ましくは、驚き動作生成装置は、音声信号からフォルマントを抽出するフォルマント抽出手段と、フォルマント抽出手段により抽出されたフォルマントに対応してロボットの口唇の開口量を制御するための口唇動作制御手段をさらに含む。
【図面の簡単な説明】
【００２０】
【図１】分析データにおける、形態素ごとの驚き発話の分布を示す図である。
【図２】分析データ中の驚き発話の驚きの度合による分布を表形式で示す図である。
【図３】分析データ中の驚き発話に伴う動作タイプの分布を示す図である。
【図４】分布データ中の各驚き表現における動作の出現度を示す図である。
【図５】分布データ中の驚き表現に伴う各動作タイプに対する動作の出現度を示す図である。
【図６】分布データ中の各形態素に関する動作タイプの分布を示す図である。
【図７】眉毛の引き上げ時間及び戻り動作の持続時間の平均及び標準偏差を驚きのレベル別に示す図である。
【図８】上半身反り動作における開始及び戻りに要する時間の平均と標準偏差とをレベル１及びレベル２の双方について示した図である。
【図９】各動作タイプについて、動作と驚き発話との間の時間差の分布を示す図である。
【図１０】アンドロイドに設けられたアクチュエータの配置を示す図である。
【図１１】アンドロイドの中立状態の表情、レベル１の驚き表情、及びレベル２の驚き表情を示す図である。
【図１２】アンドロイドの制御装置の構成を示すブロック図である。
【図１３】図１２に示す制御装置を実現するコンピュータのハードウェア構成を示すブロック図である。
【発明を実施するための形態】
【００２１】
　以下の説明及び図面では、同一の部品には同一の参照番号を付してある。したがって、それらについての詳細な説明は繰返さない。
【００２２】
　本実施の形態は、驚き発話に焦点を当て、対面対話データを分析し、驚き表現の度合又は種類と、それに伴う表情及び身体動作との関連性を分析し、その結果に基づいて人らしい驚き動作を実現した。
【００２３】
　［分析データ］
　〈マルチモーダル対話音声データベース〉
　分析には、出願人が収集したマルチモーダル対話音声データベースを用いた。このデータベースはさまざまな年代の話者による、音声、頭部のモーションキャプチャデータおよびビデオデータを含む。各対話は１０分程度で、自由会話である。データベースは発話区間の情報と書き起こしを含み、驚きが表現された発話にはラベルが付与されている。
【００２４】
　データベースには、フレーズ（アクセント句）毎に談話機能ラベルが付与され、感動詞

には、相槌、肯定、聞き返し、意外、驚き、感心、気づきなどのパラ言語ラベルが付与されている。書き起こしには感嘆詞「！」が付いているものもあり、感動詞以外のフレーズで驚きを表現したものも含まれている。そこで、「！」を含むフレーズを抽出し、明らかに驚き以外のものは除外して、驚きラベルが付与された感動詞と合わせて分析データとして用いた。分析には２８名の成人話者（男性１２名、女性１６名）のデータを用いた。このデータからおよそ６３６個の驚き発話が抽出された。

【００２５】

図１に形態素ごとの驚き発話の分布を示す。感動詞「え」が最も出現度が多く(４０％)、感動詞「あ」(２０％)と「へ」(１１％)が次いで多く出現した。

【００２６】

〈驚きの度合と種類のラベルデータ〉

各驚き発話において、驚きの度合を４段階（驚いていない、少し驚いている、驚いている、すごく驚いている）で付与するようラベル付与者に指示した。その際に、会話の文脈を考慮するため、発話の前後５秒の区間も含め、対話の相手の声も聴ける様にした。

【００２７】

驚き発話の中には、感情的・自発的なものと意図的に表現したものとが混合している。そこで、自発的か意図的かに関わらず、驚きの表現の度合を判断するようラベル付与者に指示した。驚きの表現の種類は以下の項目で付与するよう指示した。

【００２８】

１　感情的・自発的な反応

２　社会的・意図的な反応

３　対話の中に過去の驚き発話を再現・引用

意図的な驚き表現は、対話インタラクションを円滑にするため、自然に発生することが多い。しかし、驚き表現が自発的か意図的かを第三者が判断するのは、ラベル付与者の主観によるため、判断結果にばらつきが多く生じると予想される。

【００２９】

日本語母語話者４名が、各発話において驚きの度合および驚きの種類のラベルを付与した。驚きの度合は、平均を取り、０～３の指標で表現した。各度合の出現数は図２に示すとおり、レベル０：６３個、１：３６１個、２：１８７個、３：２５個となった。日常会話で最も多く出現するのは、驚き表現の度合が低いもの（レベル１）である。

【００３０】

驚いていないと判断された６３個の発話（レベル０）は除外し、残りの５７３個の発話を分析対象とした。驚き表現の種類に関しては、出現数が｛自発的：３４３個（６０％）、意図的：１８７個（３３％）、再現：４３個（７．５％）｝となった。

【００３１】

〈驚き発話に伴う動作のラベルデータ〉

驚き発話に伴う表情や身体動作に関連するラベルとして、以下の項目を使用した。

【００３２】

瞼：｛閉じている、普通に開いている、大きく見開いている｝

眉毛：｛上がっている、少し上がっている、上がっていない、顰めている｝

頭部：｛動いていない、上、下、左か右、上昇下降、傾げ、頷き、その他｝

胴体：｛動いていない、前、後ろ、上、下、左か右、傾げ、回る、その他｝

各驚き発話区間に対し、研究補助員１名がビデオを見ながら表情や動作に関連するラベルを付与した。複数の動作が含まれる場合は、複数項目の選択を可能とした。

【００３３】

図３に驚き発話で観測された動作の分布を示す。図３を参照して、眉毛を上げる動作（目を見開く動作と伴うことが多い）が最も多く、次いで頭部を上げるまたは上げてすぐ下げる動作、上半身を後ろに反る動作、及び頷き動作が多く出現した。動作が全く伴わない驚き発話は３０％で観測された。これは日常会話で驚きを表現する度に表情や動作が必ずしも伴わないことを示している。また、動作の出現度は、驚き表現の度合にも依存すると

考えられる。
【００３４】
［分析結果］
図４に、驚き表現の種類別で、驚きの度合における動作の出現度の分布を示す。全体的に驚き表現の度合が上がるにつれて、動作の出現度も上がることが分かる(「全体」)。この傾向は、感情的・自発的な驚き発話でより顕著に現れる。一方で、社会的・意図的な驚き発話および驚きの引用発話では、驚きの度合が高くても動作が必ずしも伴わないことが分かる。
【００３５】
図５に、出現度の高かった動作（眉毛、上半身、頭部）において、驚きの度合における動作の出現度の分布を示す。この結果より、眉毛を上げる動作は、驚きの度合が中および高（レベル２及び３）の場合に高くなるのに対し、上半身を後ろに反る動作は、驚きの度合が高（レベル３）の場合のみ高くなる。
【００３６】
図６に、出現度が最も多かった感動詞「え」、「あ」および「へ」における各動作の出現度の分布を示す。感動詞によって、驚き発話に伴う動作の傾向も変わることが分かる。「え」の場合は、眉毛を上げる動作が最も多く、「あ」の場合は頭を上げる動作が多かった。「へ」の場合は、頷き動作が最も多く、驚きと感心や同情の表現が伴うことが原因と考えられる。
【００３７】
以上の分析の結果、驚き表現の種類、度合、及び形態素によっても、生成すべき動作が異なることが明らかとなった。
【００３８】
さらに、驚き表現において重要なのは、各動作について、発話と自然な形で同期するよう、動作の開始及び戻りのタイミングをコントロールすることである。そのための驚き発話の分析結果について以下に説明する。
【００３９】
眉毛及び胴部の開始及び戻りに要する時間を、「え」及び「あ」という感動詞について測定した。結果を図７に示す。図７には、眉毛の引き上げと戻りの時間について、２つのレベルについて測定した結果の平均と標準偏差とを示す。レベル１の開始に要する時間（開始時間）は「開始１」、戻りは「戻り１」で示し、レベル２の開始時間を「開始２」、戻り時間を「戻り２」により示す。
【００４０】
図７から明らかな様に、いずれのレベルでも開始時間の方が戻りに要する時間より短い。開始時間の平均は２００から３００ミリ秒、戻り時間の平均は４００から５００ミリ秒である。レベル２はレベル１より多少時間が長くなっているが、動く距離が長いためである。レベル２における戻り時間の標準偏差はレベル１の戻り時間の標準偏差よりはるかに大きいが、これは眉毛が通常の位置に戻るのに１秒から２秒を要するためである。アンドロイドによる驚き動作でもこれと同様の制御をすることが望ましい。これらの制御をするための時間を、図７に示すような標準偏差に基づいて変化させる様にしてもよい。これは瞼の開大に対しても当てはまる。
【００４１】
図８は、上半身反り動作における開始及び戻りに要する時間の平均と標準偏差とをレベル１及びレベル２の双方について示したものである。上半身については、開始時間及び戻り時間がレベル１ではいずれも０．８秒程度、レベル２ではそれぞれ１．２秒及び１．５秒であった。アンドロイドによる驚き動作でもこれと同様の制御をすることが望ましい。この時間をばらつかせる様にしてもよいことは眉毛の引き上げの場合と同様である。
【００４２】
図９は各動作タイプについて、動作と驚き発話との間の時間差を示す。「開始」は動作の開始時刻と発話の開始時刻との間の差を示す。「終了」は動作の終了時刻と驚き発話の

終了時刻との間の差を示す。この結果から、眉毛、頭部、及び上半身の動作の開始時刻と、驚き発話の開始時刻との差はほぼ－０．１秒～０．１秒の間であることが分かった。即ち、通常、驚きに伴う動作は驚き発話とほぼ同期している。これに対し、終了時刻の差については、よりばらついているが大部分がプラスであることが分かった。すなわち、驚き動作の後に体が通常の中立状態の位置に戻るのは、驚き発話が終わってからということである。

【００４３】

［アンドロイドロボットにおける動作生成］

本実施の形態では、驚き発話の持続時間が所与のものであるとして、種々の驚き度合いについて種々の驚き動作を制御する場合の効果について検討した。分析の結果、発明者は、以下の４つの要件を考慮した動作生成方法を提案する。すなわち、表情制御（瞼の開大を伴う眉毛の引き上げ）、頭部動作制御（頭部のピッチ方向）、及び上半身動作制御（トルソのピッチ方向）である。

【００４４】

〈アンドロイドのアクチュエータと制御方法〉

本実施の形態で制御対象となるのは女性のアンドロイドであり、このアンドロイドを用いて驚き発話に伴う動作を評価した。アンドロイドは、遠隔の話者から送られてくる音声信号に応じて、アンドロイドの前にいる相手と対話することが想定されている。アンドロイドの発声する音声は、遠隔の話者から送られてくる音声を再生したものでもよいし、別の音声に変換したものでもよい。

【００４５】

図１０に、このアンドロイドの上体の動作を制御するアクチュエータの配置を示す。図に示す様に、このアンドロイドの上半身に設けられたアクチュエータは全部で１９個である。図１０にはアクチュエータ１～１１、１３～１９を示してある。アクチュエータ１２は舌の動きを制御するものであり、図１０には示していない。またアクチュエータ１７は肩の動きのためのものであり、本実施の形態による驚き動作制御には関与しない。すなわち、これらのうち、アクチュエータ１～１６、１８及び１９が驚き動作における制御の対象となる。

【００４６】

表情制御にはアクチュエータ１～１３が関与し、頭部動作制御にはアクチュエータ１４～１６が関与し、上半身制御にはアクチュエータ１８及び１９が関与する。すなわち、表情制御の自由度は１３、頭部動作の自由度は３、上半身制御の自由度は２である。

【００４７】

驚き動作に関与するアクチュエータは以下のとおりである。

【００４８】

上瞼制御（アクチュエータ１）
下瞼制御（アクチュエータ５）
眉毛引き上げ制御（アクチュエータ６）
口角引き上げ制御（アクチュエータ８。ほほも同時に引き上げられる。）
口角伸長制御（アクチュエータ１０）
顎部引き下げ制御（アクチュエータ１３）
頭部ピッチ制御（アクチュエータ１５）
上半身ピッチ制御（アクチュエータ１８）

全てのアクチュエータについて、コマンド値は０～２５５の範囲で与えられる。なお、以下の説明では、アクチュエータの状態（位置）については、そのコマンド値を用いて表し、「アクチュエータ値」と呼ぶことがある。

【００４９】

－眉毛及び瞼の動作制御－

驚き動作における顔の表情について、眉毛の引き上げと瞼の開大は、驚きの２つのレベルにしたがって互いに協働するよう制御される。眉毛のアクチュエータ６の位置はレベル

１については１２７、レベル２については２５５とし、上瞼及び下瞼のアクチュエータ１及び５についての目標値はレベル１についてはそれぞれ８０及び６０，レベル２についてはそれぞれ４０及び３０とする。中立状態の位置（レベル０に対応）における顔表情については、アクチュエータ６、１、５の位置はそれぞれ０、９０及び８０である。これらの値は、アンドロイドロボットの顔の表情を見ながら手動により設定した。この設定により、レベル１については少し驚いた表情、レベル２については明らかな驚きの表情が得られた。

【００５０】

図１１に、これらに対応するアンドロイドの顔の写真を示す。図１１（Ａ）は中立状態、（Ｂ）はレベル１の驚きの表情、（Ｃ）はレベル２の驚きの表情である。

【００５１】

上記した分析結果にしたがい、瞼及び眉毛のアクチュエータ１、５及び６に対しては、驚き発話の開始と同時に、各レベルに応じた値を送信し、驚き発話が終了してから０．５秒経過するまでに、滑らかに中立状態の位置に戻るようコサイン関数によりコマンドを生成する。また、上記分析に伴い、顔の表情が中立状態に戻ったときに瞬きが発生することが多いということが判明したので、本実施の形態でも、顔の表情を中立状態に戻したときに目を短時間（１００ミリ秒）閉じて中立状態に戻すことで瞬きを実現した。具体的には、アクチュエータ１及び５にコマンド２５５を送って１００ミリ秒の間目を閉じさせ、その後にそれぞれコマンド９０及び８０を送って中立状態の位置に戻す様にした。

３．１．２　上半身動作制御

上半身動作制御に関し、本実施の形態では以下の式（１）に示される様に、コサイン関数にしたがって上半身を後方に動かす。式（１）において「act[18][t]」は、アクチュエータ１８に対して時刻ｔ秒に送信されるコマンドを表す。

【００５２】

【数１】

$$act[18][t]=32+\left(upbody_{target}-32\right)*\frac{1-\cos\left(\pi\frac{t-t_{start}}{T_{onset}}\right)}{2} \qquad (1)$$

式（１）においてt_{start}は驚き動作の開始時刻、すなわち驚き発話の開始時刻であり、T_{onset}は驚き動作の開始動作の持続時間である。また$upbody_{target}$は驚き動作に伴う上半身の反り量の最大値を示す。式中の「３２」という値はアンドロイドの中立状態の姿勢におけるアクチュエータ１８の位置である。式（１）にしたがって驚き動作を行わせることにより、所与の時間T_{onset}にアンドロイドの上半身の反り量を滑らかに所与の最大値$upbody_{target}$に到達させることができる。

【００５３】

$upbody_{target}$の値は、本実施の形態では、レベル１については１６、レベル２については0である。ただし、本実施の形態では、アクチュエータ値が小さい程上半身が後ろ側に反る設定である。もちろん、これらの値は設計により、使用するアクチュエータにより、それぞれ変えることができる。

【００５４】

驚き発話の終了時刻から、上体を次の式（２）に示す様にコサイン関数にしたがって中立状態の位置に戻す。

【００５５】

【数２】

$$act[18][t]=32+\left(upbody_{end}-32\right)*\frac{1+\cos\left(\pi\frac{t-t_{end}}{T_{offset}}\right)}{2} \qquad (2)$$

式（２）において$updoby_{end}$及びt_{end}はそれぞれ、驚き発話の終了時におけるアクチュエ

ータ１８の値及び時刻である。したがって、驚き発話の発話区間がT_{onset}よりも短いと、上半身の反り量は最大値まで達しない。
【００５６】
上記した分析の結果、本実施の形態における驚き動作において、動作の開始から最大値に到達するまでの持続時間は０．８秒に設定した。驚き発話の終了時から中立状態の位置に戻るまでの時間T_{offset}は１．５秒に設定した。
【００５７】
-頭部動作制御-
頭部動作に関し、本実施の形態では、頭部のピッチを音声ピッチ（基本周波数Ｆ０）により制御する方法を採用した。実際の人間がこのような頭部動作をしているわけではないが、笑いに関する同様の研究から、このような制御をしても自然な頭部動作を実現できることが判明しているため、本実施の形態でも同様の制御を行うこととした。通常、驚いたときには、人間は高いＦ０の音声を発生して頭を上に動かして顔を上に向け、又は上に向けた後に下に向ける動作をすることが多いためである。本実施の形態では、このＦ０の値を頭部ピッチアクチュエータ１５へのコマンド値に変換するために、以下の式（３）を用いる。
【００５８】
【数３】

$$act[15][t]=128+(F0[t]-center_F0)*F0_scale \quad (3)$$

式（３）において、ｃｅｎｔｅｒ＿Ｆ０は発話者の平均Ｆ０の値（男性は約１２０Ｈｚ、女性は約２４０Ｈｚ）であって、半音単位に変換されたものであり、Ｆ０は現在のＦ０の値（半音値）であり、Ｆ０＿ｓｃａｌｅはＦ０の変化を頭部のピッチ動作にマッピングするためのスケールファクタである。本実施の形態では、Ｆ０＿ｓｃａｌｅは音声ピッチの１半音の変化が頭部のピッチ方向の回転の１度にほぼ対応する様に設定した。
【００５９】
予備実験から、驚き発話の際に上半身が後方にそったときの顔が上を向くと不自然に感じられることが分かった。また、観察により、通常の話者ではそうした場合でも顔を対話の相手方に向けていることが分かった。そのため、頭部ピッチアクチュエータ１５について、上半身の後方への反り動作の際に、逆方向の動きを加える制御を行うことにした。そのため、コマンド値は以下の式（４）により計算している。
【００６０】
【数４】

$$act[15][t]=act[15][t]-(32-act[18][t]),\quad (act[18][t]<32\text{のとき}) \quad (4)$$

-口唇動作制御-
有声で発話されない驚き表現では、顎部を下げるべきである。しかし、有声で発話される驚き表現では、口唇の動きと、それに続く顎部の動きは、ともに発話内容と同期すべきである。本実施の形態では、口唇動作は特開２０１２－１７３３８９号公報において提案された、フォルマントに基づく口唇動作制御方法を採用した。この様にすることで、種々の母音音質を持つ有声の驚き発話部分（「えー」、「あー」等）における適切な口唇形状を生成できる。この方法は母音のフォルマントを元にしているためである。顎部アクチュエータ（アクチュエータ１３）はこの様にして推定された口唇の高さを用いて制御する。
【００６１】
［構成］
以上、制御の細部について説明した本実施の形態に係る驚き動作の生成装置は以下のような構成を有する。図１２を参照して、この実施の形態に係る驚き動作生成装置２００は音声信号２０２を受けて、発話の内容、韻律、及び声質に基づいて驚き区間を検出し、その開始と終了とを少なくとも示す驚き区間信号を生成し出力するための驚き区間信号生成部２１０と、音声信号２０２からフォルマントを抽出しフォルマントの値を示す信号を出

力するフォルマント抽出部２１２と、フォルマント抽出部２１２の出力する信号に基づいて驚き区間中の口唇動作のためのアクチュエータ群（図示せず）を制御するための口唇動作制御部２１４とを含む。遠隔の話者の音声に基づいて、その話者の驚き状態を検出し、その驚き状態に応じ、実際の人間に近い自然な動きでアンドロイドの表情、頭部、及び上半身を制御する。

【００６２】

本実施の形態では、アンドロイドの状態（眉毛の位置及び瞼の開大量、頭部の位置及び上半身の状態）は、中立状態、第１のレベルの驚き状態、第２のレベルの驚き状態に制御可能である。それら状態は、各部を動作させるアクチュエータへのコマンド値をそれぞれ所定の値に制御することにより実現できる。

【００６３】

なお、以下の説明から明らかな様に、第２のレベルの驚き状態における各部の、中立状態における位置からの離間量は、第１のレベルの驚き状態における離間量よりも大きい。また、本実施の形態では、各部を第１又は第２のレベルの驚き状態から中立状態に戻すのに要する時間が、中立状態から第１又は第２のレベルの驚き状態に変化させるのに要する時間よりも長くなる様にアクチュエータを制御する。以下の実施の形態では表情、頭部、及び上半身のいずれも、第１及び第２のレベルの驚き状態とで互いに異なる動作をしている。しかし、これらがいずれもそのような動作をする必要はない。表情、頭部、及び上半身の任意の組み合わせについて、第１及び第２のレベルの驚き状態によって制御を異ならせる様にし、残りの部分については同じ制御を行う様にしてもよい。

【００６４】

フォルマント抽出部２１２としては、上述の通り特開２０１２－１７３３８９号公報に開示された口唇動作パラメータ生成装置と同様の構成を採用すれば良い。

【００６５】

驚き区間信号生成部２１０は、音声信号２０２に対して発話内容の音声認識を行い、認識結果である発話内容のテキストを出力する音声認識装置２６０と、音声信号２０２から音声の韻律及び声質特徴を抽出するための韻律・声質特徴抽出部２６２と、音声認識装置２６０から出力される発話内容のテキストと、韻律・声質特徴抽出部２６２が出力する韻律・声質特徴とを用いて発話中の驚き区間の開始及び終了位置、並びに驚きのレベル（１、２）を検出し、驚き区間信号を出力するための驚き区間検出部２６４とを含む。韻律・声質特徴抽出部２６２が抽出する特徴は、Ｆ０、規則性、及び力み等である。驚き区間検出部２６４としては、例えば特開２０１０－２１７５０２号公報に開示された発話意図情報検出の方法を採用すればよい。

【００６６】

驚き動作生成装置２００はさらに、驚き区間検出部２６４が出力する驚き区間検出信号に応答して、アンドロイドの表情を制御する信号を生成するための表情制御部２１６を含む。

【００６７】

表情制御部２１６は、驚き区間信号に応答して、その開始時刻及び終了時刻をトリガーにアンドロイドの眉毛の動作をさせるためのアクチュエータ信号を生成し、眉毛に関するアクチュエータ群２２４（アクチュエータ６）に出力するための眉毛引き上げ動作制御部２８０と、驚き区間信号に応答して、その開始時刻及び終了時刻をトリガーに、瞼に開大動作をさせるためのアクチュエータ信号を生成し、瞼の動作を制御するアクチュエータ群２２６（アクチュエータ１及び５）に出力するための瞼開大動作制御部２８２とを含む。本実施の形態では、第１のレベルの驚き状態における眉毛の引き上げ量よりも、第２のレベルの驚き状態における引き上げ量の方が大きくなる様に制御を行う。瞼の開大量も同様である。

【００６８】

驚き動作生成装置２００はさらに、驚き区間の終了時、驚き動作から中立状態の位置にアンドロイドの姿勢等が戻ったときにアンドロイドに瞬きをさせるためのアクチュエータ

信号を生成し、瞼の動作を制御するアクチュエータ群２２６に出力する瞬き制御部２１８と、驚き区間検出部２６４が出力する驚き区間信号に応答して、韻律・声質特徴抽出部２６２が出力する韻律・声質特徴と驚き区間信号により示される驚き区間の開始及び終了に基づいてそれぞれ驚き動作の開始及び終了時の頭部の動作を制御するコマンドを生成し、頭部動作に関するアクチュエータ群２２８（アクチュエータ１５）に対して出力する頭部動作制御部２２０と、驚き区間検出部２６４が出力する驚き区間信号に応答して、韻律・声質特徴抽出部２６２が出力する韻律・声質特徴と驚き区間信号により示される驚き区間の開始及び終了に基づいてそれぞれ驚き動作の開始及び終了時の上半身の動作を制御するコマンドを生成し、腰部アクチュエータ群２３０に出力する上半身動作制御部２２２とを含む。
【００６９】
本実施の形態では、頭部ピッチの変化量（動き角度）及び上半身の反り量の何れも、第１のレベルの驚き状態における値より第２のレベルの驚き状態における方が大きくなる様に制御される。なお、驚きの開始時と終了時とにおける各部の動作はできるだけ滑らかにすることが望ましい。そのために、例えば式（１）及び（２）に示すような式で位置を制御することが望ましい。
【００７０】
［動作］
遠隔から送信されてきた話者の音声信号２０２が与えられると、驚き動作生成装置２００は以下の様に動作する。音声認識装置２６０は音声信号２０２に対する音声認識を行い、発話内容のテキストを驚き区間検出部２６４に出力する。韻律・声質特徴抽出部２６２は、音声信号２０２から発話音声の韻律及び声質特徴を抽出し、驚き区間検出部２６４に与える。驚き区間検出部２６４は、発話内容のテキスト及び韻律及び声質特徴に基づいて遠隔の話者の驚き状態を反映した驚き区間と驚きレベルとを検出し、少なくともその開始時刻及び終了時刻と驚きレベルとを特定する驚き区間信号を生成して、眉毛引き上げ動作制御部２８０、瞼開大動作制御部２８２、瞬き制御部２１８、頭部動作制御部２２０及び上半身動作制御部２２２に与える。
【００７１】
眉毛引き上げ動作制御部２８０及び瞼開大動作制御部２８２は、驚き発話の開始と同時に、各驚きレベルに応じた値をアクチュエータ群２２４及び２２６にそれぞれ送信し、驚き発話が終了してから０．５秒経過するまでに、滑らかに中立状態の位置に戻るようコサイン関数によりコマンドを生成しアクチュエータ群２２４及び２２６にそれぞれ送信する。
【００７２】
瞬き制御部２１８は、驚き区間の終了時、驚き動作から中立状態の位置にアンドロイドの姿勢等が戻ったときにアンドロイドに瞬きをさせるためのアクチュエータ信号を生成し、アクチュエータ群２２６に与える。具体的には、アクチュエータ１及び５にコマンド２５５を送って１００ミリ秒の間、瞼を閉じさせ、その後にそれぞれコマンド９０及び８０を送って瞼を中立状態の位置に戻すことによりアンドロイドに瞬きを行わせる。
【００７３】
頭部動作制御部２２０は、驚き区間信号に応答して、韻律・声質特徴抽出部２６２からの韻律・声質特徴を用いて頭部の動作を制御する。具体的には、頭部動作制御部２２０は、驚き区間の開始時には、韻律・声質特徴抽出部２６２から与えられるＦ０を式（３）及び式（４）を用いて頭部動作のためのコマンド値を算出し、アクチュエータ１５に対して出力する。ただし式（４）による変換を行うのは、上半身動作制御部２２２から上半身の動作のためのアクチュエータ群２３０（アクチュエータ１８）に与えられるコマンドが３２より小さいときだけである。
【００７４】
上半身動作制御部２２２は、驚き区間信号に応答し、驚き区間の開始時には式（１）にしたがって驚きレベルに応じたコマンドを生成しアクチュエータ群２３０に与える。また

上半身動作制御部２２２は、驚き区間の終了時には式（２）にしたがって驚きレベルに応じたコマンドを生成しアクチュエータ群２３０に与えて上半身を中立状態の位置に戻す。
【００７５】
一方、フォルマント抽出部２１２は驚き区間信号生成部２１０からフォルマントを抽出し口唇動作制御部２１４に与える。口唇動作制御部２１４はフォルマントに基づく口唇動作制御方法にしたがい、種々の母音音質を持つ有声の驚き発話部分に応じた適切な口唇形状を生成するためのコマンドを生成し、口唇動作を行うアクチュエータ１３に与えて口唇及び顎部を動作させる。口唇動作制御部２１４は驚き信号とは無関係に動作するが、この様にフォルマントに基づいて口唇動作を制御しても、音声の声質等に応じて驚き区間でも適切な口唇動作を実現できる。
【００７６】
［コンピュータによる実現］
図１３は、アンドロイドを制御する装置の典型例の構成を示す。この実施の形態では、制御装置６３０はコンピュータ６４０からなり、このコンピュータ６４０をアンドロイドの各部と接続することでアンドロイドの動作を制御する。図１３を参照して、制御装置６３０は、メモリポート６５２及び入出力インターフェイス（入出力Ｉ／Ｆ）６５０を有するコンピュータ６４０と、いずれもコンピュータ６４０に接続されたキーボード６４６と、マウス６４８と、モニタ６４２とを含む。コンピュータ６４０は、入出力Ｉ／Ｆ６５０を介してアンドロイドの各部と接続されている。
【００７７】
コンピュータ６４０はさらに、ＣＰＵ（中央処理装置）６５６と、ＣＰＵ６５６、メモリポート６５２及び入出力Ｉ／Ｆ６５０に接続されたバス６６６と、起動プログラム等を記憶する読出専用メモリ（ＲＯＭ）６５８と、バス６６６に接続され、上記頷き生成装置１５０の各部の機能を実現するプログラム命令、システムプログラム及び作業データ等をプログラムの実行時に記憶するランダムアクセスメモリ（ＲＡＭ）６６０と、ハードディスク６５４を含む。コンピュータ６４０はさらに、他端末との通信を可能とするネットワーク６６８への接続を提供するネットワークインターフェイス（Ｉ／Ｆ）６４４を含む。
【００７８】
制御装置６３０を上記した実施の形態に係るアンドロイドの制御装置として機能させるためのコンピュータプログラムは、メモリポート６５２に装着されるリムーバブルメモリ６６４、又は入出力Ｉ／Ｆ６５０に接続される図示しない外部記憶装置に記憶され、さらにハードディスク６５４に転送される。又は、プログラムはネットワーク６６８を通じてコンピュータ６４０に送信されハードディスク６５４に記憶されてもよい。プログラムは実行の際にＲＡＭ６６０にロードされる。図示しない外部記憶装置から、リムーバブルメモリ６６４から又はネットワーク６６８を介して、直接にＲＡＭ６６０にプログラムをロードしてもよい。
【００７９】
このプログラムは、コンピュータ６４０を、上記実施の形態に係るアンドロイドの制御装置の各機能部として機能させるための複数の命令からなる命令列を含む。コンピュータ６４０にこの動作を行わせるのに必要な基本的機能のいくつかはコンピュータ６４０上で動作するオペレーティングシステム若しくはサードパーティのプログラム又はコンピュータ６４０にインストールされる、ダイナミックリンク可能な各種プログラミングツールキット又はプログラムライブラリにより提供される。したがって、このプログラム自体はこの実施の形態のシステム、装置及び方法を実現するのに必要な機能全てを必ずしも含まなくてよい。このプログラムは、命令のうち、所望の結果が得られる様に制御されたやり方で適切な機能又はプログラミングツールキット又はプログラムライブラリ内の適切なプログラムを実行時に動的に呼出すことにより、上記したシステム、装置又は方法としての機能を実現する命令のみを含んでいればよい。もちろん、独立したプログラムのみで必要な機能を全て提供してもよい。
【００８０】

上記実施の形態では、図１２に示す驚き区間検出部２６４にルールベースのものを用いている。しかし本発明はそのような実施の形態には限定されない。ルールベースの検出部に代えて、ＳＶＭ、ナイーブベイズ、ロジスティック回帰、ＤＮＮ、ＲＮＮ、ＣＮＮ等からなる、機械学習による判定器を用いても良い。

【００８１】

さらに、上記実施の形態では、驚き区間信号生成部２１０による驚き区間の検出を行っている。これは、遠隔の話者の音声に応じて、その話者の驚き状態をアンドロイドに反映させる必要があるためである。本発明はそのような実施の形態には限定されない。例えば、アンドロイドを所定のシナリオにしたがって動作させるような場合には、驚き区間は既知となる。したがって、驚き区間信号生成部２１０は不要であり、シナリオにしたがって、驚き動作を行うべき区間を示す信号を眉毛引き上げ動作制御部２８０、瞼開大動作制御部２８２、瞬き制御部２１８、頭部動作制御部２２０及び上半身動作制御部２２２に与える様にすればよい。この場合には、フォルマント抽出部２１２も不要で、合成出力すべき音声のターゲットからフォルマントを求めればよい。

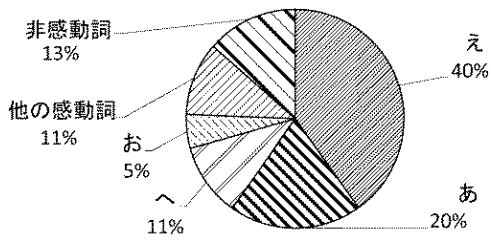
【００８２】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内での全ての変更を含む。

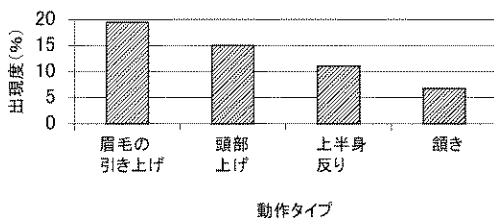
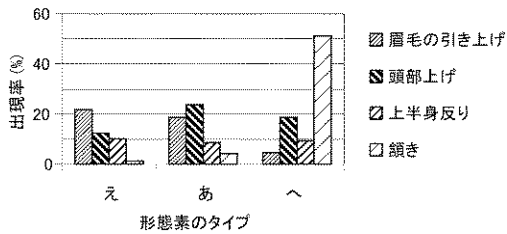
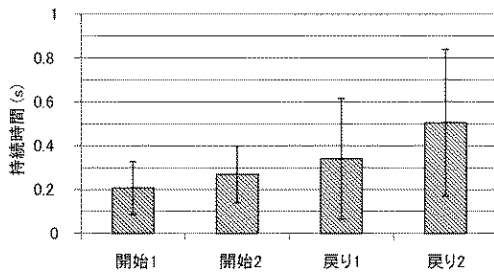
【符号の説明】

【００８３】

２００　驚き動作生成装置
２０２　音声信号
２１０　驚き区間信号生成部
２１２　フォルマント抽出部
２１４　口唇動作制御部
２１６　表情制御部
２１８　瞬き制御部
２２０　頭部動作制御部
２２２　上半身動作制御部
２２４、２２６、２２８、２３０　アクチュエータ群
２６０　音声認識装置
２６２　韻律・声質特徴抽出部
２６４　驚き区間検出部
２８０　眉毛引き上げ動作制御部
２８２　瞼開大動作制御部
６３０　制御装置

【図１】

【図２】

知覚上の驚き度合い	0	1	2	3
発話数	63	361	187	25

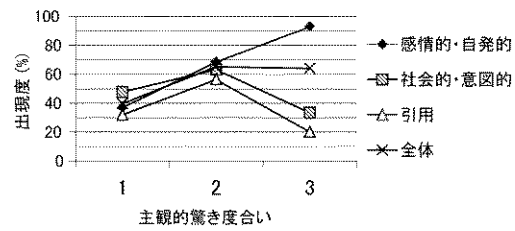
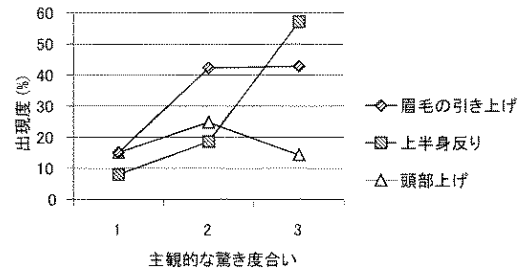
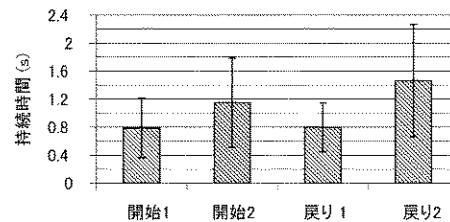
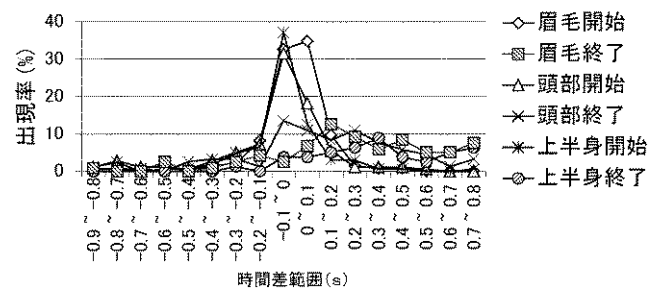
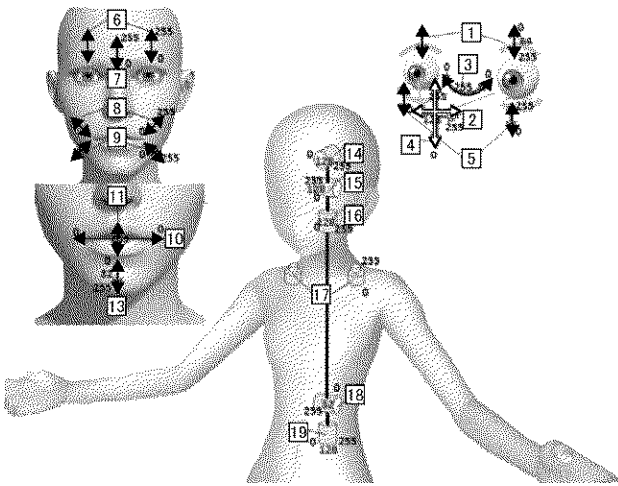
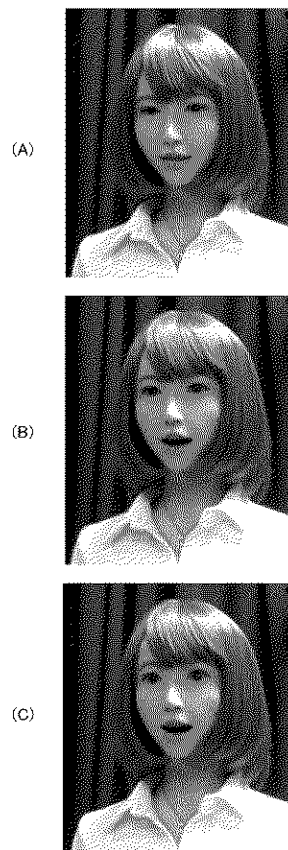
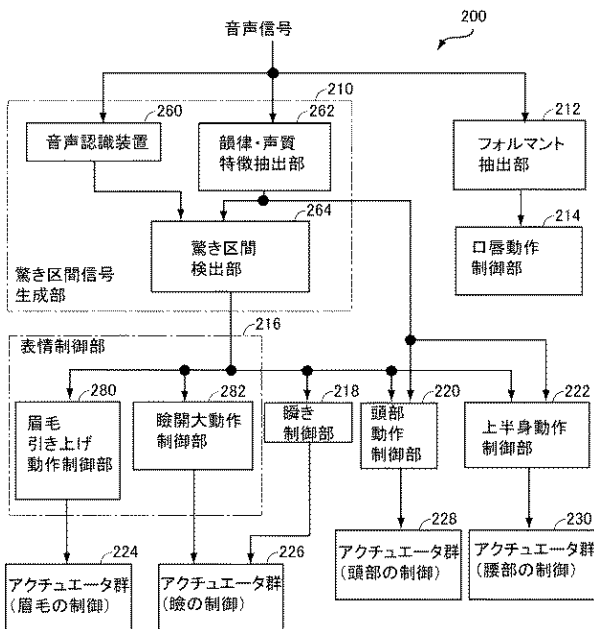
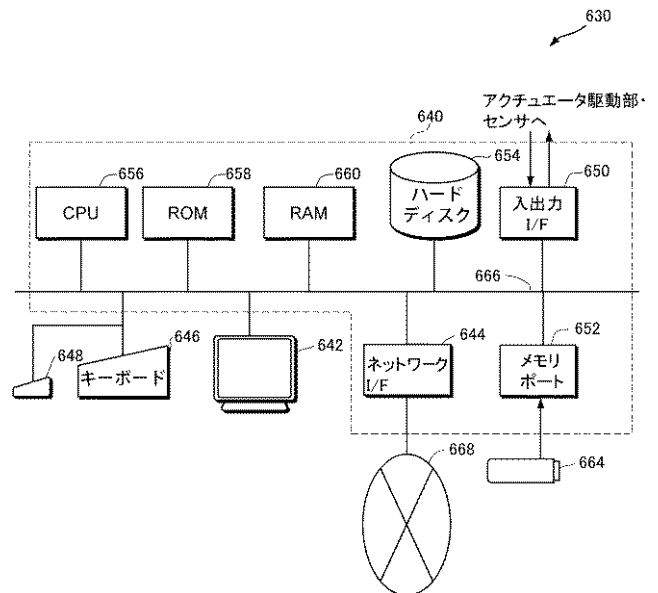
【図３】

【図４】

【図５】

【図６】

【図７】

【図８】

【図９】

【図１０】

【図１１】

【図１２】

【図１３】

フロントページの続き

(51)Int.Cl.			ＦＩ			テーマコード（参考）
Ｇ１０Ｌ	13/00	(2006.01)	Ｇ１０Ｌ	13/00	１００Ａ	
Ｇ１０Ｌ	15/00	(2013.01)	Ｇ１０Ｌ	15/00	２００Ｈ	
Ａ６３Ｈ	13/04	(2006.01)	Ａ６３Ｈ	13/04	Ｊ	
Ａ６３Ｈ	11/00	(2006.01)	Ａ６３Ｈ	11/00	Ｚ	